

J-5574 U.S. PTO
96/10/98

Please type a plus sign (+) inside this box →

PTO/SB/05 (12/97)
Approved for use through 09/30/00. OMB 0651-0032
Patent and Trademark Office: U.S. DEPARTMENT OF COMMERCE

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

UTILITY PATENT APPLICATION TRANSMITTAL

(Only for new nonprovisional applications under 37 CFR 1.53(b))

Attorney Docket No.	CIS-032-B	Total Pages	29
<i>First Named Inventor or Application Identifier</i>			
Earl Cohen			
Express Mail Label No.	EE143637897US		

APPLICATION ELEMENTS

See MPEP chapter 600 concerning utility patent application contents.

- | | |
|---|--|
| <p>1. <input type="checkbox"/> Fee Transmittal Form
<i>(Submit an original, and a duplicate for fee processing)</i></p> <p>2. <input checked="" type="checkbox"/> Specification [Total Pages 23]
<i>(preferred arrangement set forth below)</i></p> <ul style="list-style-type: none"> ✓ - Descriptive title of the Invention - Cross References to Related Applications - Statement Regarding Fed sponsored R & D - Reference to Microfiche Appendix <p>✓ - Background of the Invention</p> <p>✓ - Brief Summary of the Invention</p> <p>✓ - Brief Description of the Drawings <i>(if filed)</i></p> <p>✓ - Detailed Description</p> <p>✓ - Claim(s)</p> <p>✓ - Abstract of the Disclosure</p> <p>3. <input checked="" type="checkbox"/> Drawing(s) <i>(35 USC 113)</i> [Total Sheets 4]</p> <p>4. Oath or Declaration [Total Pages]</p> <ul style="list-style-type: none"> a. <input type="checkbox"/> Newly executed (original or copy) b. <input type="checkbox"/> Copy from a prior application (37 CFR 1.63(d))
<i>(for continuation/divisional with Box 17 completed)</i>
<i>[Note Box 5 below]</i> <p>i. <input type="checkbox"/> DELETION OF INVENTOR(S)
Signed statement attached deleting
inventor(s) named in the prior application,
see 37 CFR 1.63(d)(2) and 1.33(b).</p> <p>5. <input type="checkbox"/> Incorporation By Reference <i>(useable if Box 4b is checked)</i>
The entire disclosure of the prior application, from which a
copy of the oath or declaration is supplied under Box 4b,
is considered as being part of the disclosure of the
accompanying application and is hereby incorporated by
reference therein.</p> | <p>6. <input type="checkbox"/> Microfiche Computer Program <i>(Appendix)</i></p> <p>7. Nucleotide and/or Amino Acid Sequence Submission
<i>(if applicable, all necessary)</i></p> <ul style="list-style-type: none"> a. <input type="checkbox"/> Computer Readable Copy b. <input type="checkbox"/> Paper Copy <i>(identical to computer copy)</i> c. <input type="checkbox"/> Statement verifying identity of above copies |
|---|--|

ACCOMPANYING APPLICATION PARTS

- | |
|--|
| <p>8. <input type="checkbox"/> Assignment Papers (cover sheet & document(s))</p> <p>9. <input type="checkbox"/> 37 CFR 3.73(b) Statement <input type="checkbox"/> Power of Attorney
<i>(when there is an assignee)</i></p> <p>10. <input type="checkbox"/> English Translation Document <i>(if applicable)</i></p> <p>11. <input type="checkbox"/> Information Disclosure Statement (IDS)/PTO-1449 <input type="checkbox"/> Copies of IDS Citations</p> <p>12. <input type="checkbox"/> Preliminary Amendment</p> <p>13. <input checked="" type="checkbox"/> Return Receipt Postcard (MPEP 503)
<i>(Should be specifically itemized)</i></p> <p>14. <input type="checkbox"/> Small Entity <input type="checkbox"/> Statement filed in prior application, Statement(s) <input type="checkbox"/> Status still proper and desired</p> <p>15. <input type="checkbox"/> Certified Copy of Priority Document(s)
<i>(if foreign priority is claimed)</i></p> <p>16. <input checked="" type="checkbox"/> Other: <i>Cert. of Express Mailing.....</i></p> |
|--|

17. If a CONTINUING APPLICATION, check appropriate box and supply the requisite information:

Continuation Divisional Continuation-in-part (CIP) of prior application No: _____

18. CORRESPONDENCE ADDRESS

Customer Number or Bar Code Label

(Insert Customer No. or Attach bar code label here)

or Correspondence address below

NAME	Steven A. Swernofsky				
	The Law Offices of Steven A. Swernofsky				
ADDRESS	Post Office Box 390013				
	Mountain View		STATE	CA	ZIP CODE
COUNTRY	USA	TELEPHONE	(650) 947-0700		FAX
					(650) 947-8438

Burden Hour Statement: This form is estimated to take 0.2 hours to complete. Time will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, Patent and Trademark Office, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Assistant Commissioner for Patents, Box Patent Application, Washington, DC 20231.

This application is submitted in the name of the following inventor:

2

3 Inventor Citizenship Residence Address

4 Cohen, Earl United States 43750 Cameron Hills Drive

5 Fremont, CA 94539

6

The assignee is Cisco Technology, Inc., a California corporation having an office at 170 West Tasman Drive, San Jose, California 95124.

9

10

Title of the Invention

1

12

Route/Service Processor Scalability Via Flow-Based Distribution of Traffic

12

三

Background of the Invention

5

16 1. *Field of the Invention*

17

18 The present invention relates to architecture for a router. More particularly,
19 the present invention relates to architecture for a router with scalable processing power.

20

1 2. *The Prior Art*

2

3 Computers generate and utilize large amounts of data. Different compo-
4 nents in a computer network, for example, several different computers each with an
5 internet connection or “uplink”, need to be linked together to allow for the transfer of
6 data among the different components in the network. Routers perform this function.
7 Routers take data input from one component in a network and ensure that it is properly
8 transferred to another component in the network.

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

61

62

63

64

65

66

67

68

69

70

71

72

73

74

75

76

77

78

79

80

81

82

83

84

85

86

87

88

89

90

91

92

93

94

95

96

97

98

99

100

101

102

103

104

105

106

107

108

109

110

111

112

113

114

115

116

117

118

119

120

121

122

123

124

125

126

127

128

129

130

131

132

133

134

135

136

137

138

139

140

141

142

143

144

145

146

147

148

149

150

151

152

153

154

155

156

157

158

159

160

161

162

163

164

165

166

167

168

169

170

171

172

173

174

175

176

177

178

179

180

181

182

183

184

185

186

187

188

189

190

191

192

193

194

195

196

197

198

199

200

201

202

203

204

205

206

207

208

209

210

211

212

213

214

215

216

217

218

219

220

221

222

223

224

225

226

227

228

229

230

231

232

233

234

235

236

237

238

239

240

241

242

243

244

245

246

247

248

249

250

251

252

253

254

255

256

257

258

259

260

261

262

263

264

265

266

267

268

269

270

271

272

273

274

275

276

277

278

279

280

281

282

283

284

285

286

287

288

289

290

291

292

293

294

295

296

297

298

299

300

301

302

303

304

305

306

307

308

309

310

311

312

313

314

315

316

317

318

319

320

321

322

323

324

325

326

327

328

329

330

331

332

333

334

335

336

337

338

339

340

341

342

343

344

345

346

347

348

349

350

351

352

353

354

355

356

357

358

359

360

361

1 particular destination. Packets within the same flow will have the same
2 source/destination address, the same source/destination port, and the same protocol. It is
3 required to maintain the packet ordering within a flow.

4

5 Speed and accuracy are two important features of routers. If the data trans-
6 fer rate, or throughput rate, is not rapid enough, the end user of the required data must
7 wait for it to arrive, wasting valuable processing time. Also, input and output queues
8 waiting to transfer data may overflow if information is being added more rapidly than it
9 is being removed. Accuracy in transfer is also important to ensure that data arrives at its
10 proper destination in the proper order.

11

12 Routers can also provide services as they perform the data transfer or
13 switching. Some of these services involve gathering information about the data being
14 transferred or performing some other processing function upon the data as it is being
15 transferred from one component in the computer network to another. Some of the more
16 recent service developments involve “touching”, or performing some processing func-
17 tion, on the data in most or every packet as it is transferred. Such “high touch” services
18 can require the router system to have a great deal of processing power. New services are
19 being developed all the time, and thus the amount of processing that performing services
20 requires is unbounded.

21

1 Current routing systems generally operate on a “one processor per line
2 card” model. A line card is a group of components in a computer network such as a
3 group of modems or an internet connection which sends and receives data. In a “one
4 processor per line card” model, all of the processing for a single line, or a group of
5 related lines, is done by a dedicated processor. This model works acceptably well if the
6 line’s rate of data input, or input interface rate, is low enough so that the processor does
7 not become a bottleneck in the system. With low enough interface rates, one processor
8 can handle multiple lines.

9
10 However, the line’s interface rate may increase as higher-speed computers
11 or internet connections are added. Furthermore, additional router services may become
12 desirable. Eventually, the dedicated processor may no longer be able to keep up with the
13 processing requirements of the line and the throughput rate will slow down. If the
14 throughput rate becomes unacceptably slow, the only solution is generally to replace the
15 dedicated processor with a more powerful machine. This is an expensive solution. In
16 certain cases, for example a high-speed internet line or “uplink”, a processor may not
17 even be available that can keep up with the interface rate of the high-speed line. Current
18 internet uplink rates such as Gigabit Ethernet (1 gigabit per second) are already becoming
19 difficult to handle with a single dedicated processor.

1 able uplink connections, and which would allow processors to be added as more proc-
2 essing power became necessary, without requiring the entire system's future hardware
3 needs to be purchased initially. It would also be advantageous to maintain packet order-
4 ing within each flow while still flexibly routing among different processors as necessary.

5

6 Summary of the Invention

7

8 The invention provides a system and a method for a router architecture that
9 is scalable, that is, as more processing power is desired, more individual processors can
10 be added. The data flow from each line can be distributed among all of the processors in
11 the system. As desired services are added, increasing the amount of "touch" or process-
12 ing performed on the packets in the system, more processors can be added to carry the
13 increased load. The router architecture is also able to distribute the high interface rate of
14 an uplink connection in the same manner.

15

16 In a preferred embodiment, packets are allocated to processors in a manner
17 that allows the original order of data packets within the same flow to be maintained. The
18 system uses a hash function to distribute the flows, making sure that packets within the
19 same flow are sent to the same processor so that the original packet order in each flow is
20 maintained. Different flows may be sent to different processors.

21

1 Brief Description of the Drawings

2

3 Figure 1 shows a logical block diagram of the router architecture of the pre-
4 sent invention.

5

6 Figure 2 shows a sample flow diagram for a data packet.

7

8 Figure 3a shows an example port adapter design for an Ethernet interface.

9

10 Figure 3b shows an example port adapter design for a channelized T3 line
11 (CT3) interface.

12

13 Figure 4 shows an example Data Compiler.

14

15 Figure 5 shows an example packet digest format.

16

17 Detailed Description of a Preferred Embodiment

18

19 In the following description, a preferred embodiment of the invention is
20 described with regard to preferred process steps and data structures. However, those
21 skilled in the art would recognize, after perusal of this application, that embodiments of
22 the invention may be implemented using a set of general purpose computers operating

1 under program control, and that modification of a set of general purpose computers to
2 implement the process steps and data structures described herein would not require undue
3 invention.

4

5 The present invention provides a router architecture capable of implement-
6 ing a wide variety of services while balancing the router system load among a number of
7 processors. The router architecture of the present invention is also capable of distributing
8 the load from a single uplink among the multiple processors. The ability to distribute the
9 router system load among all of the processors in the system makes the system through-
10 put scalable; the throughput rate increases with each additional processor. Because there
11 is no cooperation problem among the processors and each is acting independently and
12 preferably on separate flows, the system throughput is expected to scale approximately
13 linearly with the number of processors.

14

15 The use of multiple processors which are redundant also provides added
16 reliability. If one processor fails, the processing tasks can be allocated among the
17 remaining processors.

18

19 FIG. 1 shows a logical block diagram of the components in the router sys-
20 tem. FIG. 1 is intended to show the interconnections and resulting scalability of the
21 router system architecture, and is not meant to be an accurate representation of the num-
22 ber of the various components in the router system in any particular embodiment. Indi-

1 individual data interface components 10 are connected to a set of Port Adapters ("PA's) 20,
2 which are in turn connected to a set of Data Compilers (DCs) 30. The DCs 30, an
3 Uplink 60, and a set of Route Processing Engines (RPEs) 50 are all interconnected
4 through an Xbar 40. The Uplink 60 is connected to an external network such as the
5 internet 70. Data transfer originates at either the individual data interface components 10
6 or the internet 70 and travels through the router system to the Xbar 40, where it is sent on
7 to be processed by a selected one of the RPEs 50. Once processing is complete, the data
8 is transferred back through the Xbar 40 to its final destination, either the internet 70 or an
9 individual data interface component 10.

5

40

51

52

53

54

55

56

57

58

59

60

61

62

63

64

65

66

67

68

69

70

71

72

73

74

75

76

77

78

79

80

81

82

83

84

85

86

87

88

89

90

91

92

93

94

95

96

97

98

99

100

101

102

103

104

105

106

107

108

109

110

111

112

113

114

115

116

117

118

119

120

121

122

123

124

125

126

127

128

129

130

131

132

133

134

135

136

137

138

139

140

141

142

143

144

145

146

147

148

149

150

151

152

153

154

155

156

157

158

159

160

161

162

163

164

165

166

167

168

169

170

171

172

173

174

175

176

177

178

179

180

181

182

183

184

185

186

187

188

189

190

191

192

193

194

195

196

197

198

199

200

201

202

203

204

205

206

207

208

209

210

211

212

213

11 individual data interface components 10 are connected to a set of Port Adapters ("PA's) 20,
12 which are in turn connected to a set of Data Compilers (DCs) 30. The DCs 30, an
13 Uplink 60, and a set of Route Processing Engines (RPEs) 50 are all interconnected
14 through an Xbar 40. The Uplink 60 is connected to an external network such as the
15 internet 70. Data transfer originates at either the individual data interface components 10
16 or the internet 70 and travels through the router system to the Xbar 40, where it is sent on
17 to be processed by a selected one of the RPEs 50. Once processing is complete, the data
18 is transferred back through the Xbar 40 to its final destination, either the internet 70 or an
19 individual data interface component 10.

20

21

1 In step 110, the data is first sent to the PA 20. A single PA 20 is capable of
2 handling the throughput from multiple modems or other individual data interface compo-
3 nents 10 since such components send data at a relatively slow and intermittent rate. The
4 PA 20 bundles this data it has received from multiple modems or other interface lines
5 into packets. Each packet is represented as a series of descriptors in a format compatible
6 with the PA 20. Network interface components, such as many Ethernet transceivers and
7 HDLC channel managers, use a descriptor approach to transfer packet data to a local
8 memory, possibly with the help of a local microprocessor. Each descriptor effectively
9 indicates a memory block (called a particle) and a size, as well as control information
10 such as whether this is the first or last descriptor in a packet. These packets are the main
11 units of data sent across computer network systems. The creation and use of data packets
12 are well known in the art of computer networking systems.

13
14 Various types of PAs are available to convert between different user inter-
15 faces and a generic PCI bus. The PCI (Peripheral Component Interconnect) bus is known
16 in both the computer and networking industries. FIG. 3a shows a representative PA de-
17 sign for an Ethernet interface. FIG. 3b shows a representative PA design for a channel-
18 ized T3 line (CT3) interface. Other user interfaces such as OC-3 (a fiber-optic standard)
19 may also be converted using a PA.

20
21 In step 112, the data packets from the PA 20 are sent to a DC 30. A DC
22 provides an interface between the internal fabric in the router and a generic PCI bus

1 which can connect to a PA. Figure 4 shows a representative Data Compiler. The DC 30
2 further prepares the packets for transfer along the routing system by putting them into
3 packet digest form, and then forwards the packets along to the Xbar 40. A packet digest
4 contains both a header and a payload. The DC 30 will further “packetize” the incoming
5 packets by converting them from the particles to a consecutive block of memory, forming
6 the “payload” for a packet digest. The DC 30 will also add a packet digest header.

7

8 The packet digest header contains information about the packet to aid in the
9 prompt forwarding and processing of the packet, such as—the Xbar 40 source port num-
10 ber, the size of the packet digest header and payload, the destination RPE 50, the packet
11 priority, the protocol type, and the class of service required.

12

13 Figure 5 shows an example packet digest format. In a preferred embodiment, the
14 packet digest includes the following fields:

15

- 16 o a CMD field to control features of the packet digest (such as the TIMESTAMP field, as
17 described below), and to indicate a type of packet digest (where there is more than one
18 type, such as in alternative embodiments);
- 19
- 20 o a SRC_PORT field to indicate a source ID of the transmitting entity;
- 21
- 22 o a PKT_OFF field and a NEW_PKT_OFF field to indicate an number of bytes of padding
23 between the packet digest header and the packet digest payload (where the PKT_OFF in-

1 dicates a total size of the packet digest header plus any padding, while the
2 NEW_PKT_OFF field indicates a new value for PKT_OFF);

3
4 o a PDH_SIZE field to indicate the size of the packet digest header;

5
6 o a DEST_OFFSET field to indicate an offset of the destination IP address in the packet digest
7 header;

8
9 o a “FIB Leaf” field to indicate a pointer to the leaf in a FIB lookup m-trie for the destina-
10 tion IP address;

11
12 o a DEST_MASK field to indicate which Xbar ports are intended as destinations for the
13 packet (where a selected bit <39> of the DEST_MASK field preferably indicates a multi-
14 cast packet);

15
16 o a QUEUE field to indicate to which hardware output queue at the particular output
17 interface to direct the packet;

18
19 o an INTERFACE field to indicate to which hardware output interface to direct the
20 packet;

21
22 o a TIMESTAMP field to indicate a timestamp or other data to be written by the in-
23 bound Xbar interface (where a selected bit CMD <0> indicates that the TIME-
24 STAMP field should be written to); and

1

2 o a set of additional words reserved for software classification data.

3

4 In order to determine the destination RPE 50, the DC 30 includes a classifi-
5 cation engine coupled to the crossbar interface to determine which RPE 50 will receive
6 each packet. Packets are not distributed randomly among RPEs, so as to maintain the
7 original ordering of packets within the same flow. The classification engine distributes
8 packets so that packets within the same flow will be sent to the same RPE, but packets in
9 different flows may be sent to different RPEs and possibly reordered as determined by
10 RPE resources.

11

12 The classification engine analyzes the router system traffic and distributes
13 flows to multiple processors. The distribution is partly table-driven, and is thus flexible
14 and easily changeable, even dynamically. In a preferred embodiment, the classification
15 uses a hash function of packet flow information to select a specific RPE 50 to handle
16 processing for that packet flow. The hash function distributes packets evenly among the
17 processors in response to flow information such as the source/destination address, the
18 source/destination port, and the protocol. The hash function can operate using any in-
19 formation that will allow for flow preservation.

20

1 In a preferred embodiment, the hash function performs an XOR (exclusive
2 OR) logical function of several bytes of fields from the IP header of the packet, including
3 the following fields:
4
5 o a source IP address (four bytes);
6
7 o a destination IP address (four bytes); and
8
9 o a protocol type value (one byte).
10
11
12
13

14
15 The XOR logical function is performed for all nine bytes, providing a single byte output value, which can be used to index into a 256-entry table for selecting the
16 RPE 50.

17
18
19
20
21
22 Once the DC 30 has determined the packet's destination RPE 50 and the
packet header and payload are ready to be sent out, the packet (now in packet digest
form) is placed in an input queue to be stored until it is sent to the selected RPE 50. Note
that the amount of processing performed by the DC 30 on each packet is bounded; the
DC 30 only does a fixed amount of work per packet. There is no cost in terms of DC 30
processing for adding additional services, as all such additional processing will be per-
formed by the RPEs 50. Thus, the DC 30 is prevented from becoming a system bottle-
neck as more services are added. In the router system of the present invention, DC 30

1 input queues should only overflow if there are not enough RPEs 50 to support the re-
2 quested level of services at the incoming data rates; if input queues overflow, it is always
3 possible to add more RPEs 50.

4

5 In step 114, packets are sent from the DC 30 to the Xbar 40. The Xbar 40
6 provides the interconnection between the data origination component ports, the router
7 system processor ports, and the data destination component ports. In one data transfer,
8 the data packet will cross through the Xbar 40 twice (step 114 and step 118 in FIG. 2).
9 The Xbar 40 receives an incoming packet and transfers it to the selected RPE, and then
10 the Xbar 40 transfers the packet back to its destination port. The Xbar 40 must provide
11 enough bandwidth to handle the data interface rates of both the DCs 30 and the
12 Uplink 60. Xbar interconnections are known in the art of computer and router systems.

13

14 The Xbar 40 includes a number of ports; each port is a pair of uni-
15 directional links. The Xbar 40 implements a simple arbitration scheme between sending
16 and receiving components, ignoring whether the destination port has the resources to
17 handle the packet or not. Packets can be dropped at any point if there are not sufficient
18 resources to handle them. Software must ensure that there are sufficient resources avail-
19 able at a destination port, for example, memory space at the output to queue the packet,
20 before sending a packet over the Xbar 40. Output queuing protocols determine what oc-
21 curs when an output queue overflows; output queue management is performed at the
22 DC 30.

1

2 The Xbar 40 includes an arbiter — ports forward their requests to the arbit-
3 ter, and grants are sent back. The Xbar 40 arbiter uses a windowing scheme in which
4 each port can make requests for multiple destinations at the same time; the arbiter at-
5 tempts to find the best fit between requesting ports and destinations. The Xbar 40 can
6 also support multicast — a port can request that a packet be sent to multiple destinations.
7 In the preferred embodiment, the Xbar 40 waits for all ports to be free, and then sends the
8 multicast request all at once. Numerous other Xbar multicast schemes are well known in
9 the art.

10
11

12 The Xbar 40 in the preferred embodiment discussed herein is packet-based,
13 but the Xbar 40 could be designed to be cell-based and still fall within the inventive con-
14 cepts disclosed herein. It is also possible to provide multiple Xbars or a bus that per-
15 forms the same function as the Xbar. The methods discussed herein for the implemen-
16 tation of such an Xbar device are by no means limiting. Alternate methods for the imple-
17 mentation of such an Xbar device will readily suggest themselves to those of ordinary
18 skill in the art.

19

20 In step 116, the Xbar 40 transfers the packet to the destination RPE 50.
21 There are multiple RPEs 50 connected to the Xbar 40 that provide both switching and
22 services, including the high-touch type of services that may require a great deal of proc-
essing power. Such services may also be added later to the system as more services are

1 developed and become available. Additional RPEs 50 can be added to the router system
2 to provide additional processing power; this provides the scalability feature of the present
3 invention.

4

5 An RPE 50 receives a packet digest with a packet digest header containing
6 classification information provided by the DC 30. Based on the classification informa-
7 tion, the packet is put into one of a small number of input queues on the RPE 50. The
8 intent of the RPE 50 queuing is to provide the capability for handling both high- and low-
9 priority packets. Packets are processed from among the RPE's input queues by the
10 RPE 50 processor in whatever order it sees fit.

11

12

The RPE 50 processor will perform various tasks on the packet, including
13 switching, tag application/update, access list processing (filtering), and all other router
14 service functions. Some of these operations result in the RPE 50 modifying the packet
15 digest, others only update RPE-internal data structures. When the RPE 50 processor is
16 done with a packet, the RPE 50 will know the packet's final destination. The example
17 destination in FIG. 2 is the internet 70 via the Uplink connection 60. However, it would
18 be evident to one of ordinary skill in the art that the packet's destination could also be an
19 individual data interface component 10 via a DC 30 and a PA 20.

20

21

22

In step 118, the packet is put back on the Xbar 40 from the RPE 50. In
step 120, the Xbar 40 transmits the packet to the Uplink 60. The Uplink 60 is a high-

1 bandwidth PA with its own DC. Like a DC, the Uplink 60 is connected to the Xbar 40
2 via an Xbar link. However, in order to meet its bandwidth requirements, an Uplink 60
3 may require multiple Xbar ports. The Xbar 40 can support multiple Uplink 60
4 connections.

5

6 In step 122, the Uplink 60 transfers the data to the internet 70, completing
7 the data transfer from the individual data interface component 10 to the internet 70.

8

9 The Uplink(s) 60 can also send data to the Xbar 40 for transmission to an
10 RPE 50 and eventual transfer to an individual data interface component 10. The
11 Uplink(s) 60's traffic from the internet 70 is also distributed via the Xbar 40 to the vari-
12 ous RPEs 50 using a hash function. In this way, the router system can support a plurality
13 of Uplink(s), each of arbitrary speed.

14

15 Various alternate embodiments of the present invention have also been
16 conceived, and would be clear to those skilled in the art after reviewing this application.
17 For example, in alternative embodiments, the DC and a specific PA type can be inte-
18 grated together to produce a more cost-effective, though less easily adaptable, interface.

19

20 Other alternative embodiments include performing some services of a
21 bounded nature on the DCs, as is currently done with output queuing. For example, the
22 DCs could accumulate statistics and do flow-based accounting on each packet. In this

1 way, the DCs could handle some portion of the known processing load leaving un-
2 bounded and future services to the RPEs.

3

4 Yet other alternative embodiments include adding specialized processing
5 engines connected to the Xbar. Packets can be sent through these specialized engines
6 either before or after or instead of an RPE to perform services such as compression/
7 decompression, encryption, or routing. The classification engine on a DC could
8 determine the type of service required by a packet and route it appropriately.

9

10

11

12

13

14

While embodiments and applications of this invention have been shown and described, it would be apparent to those skilled in the art that many more modifications than those mentioned above are possible without departing from the inventive concepts herein. The invention, therefore, is not to be restricted except in the spirit of the appended claims.

1 Claims

2

3 1. A scalable routing system for distributing packets in a network, in-
4 cluding
5 a plurality of data compilers;
6 a plurality of port adapters connected to said data compilers;
7 a plurality of route processing engines; and
8 at least one structure connecting said plurality of route processing engines
9 to said data compilers.

10

11

12 2. The scalable routing system of claim 1, including at least one uplink
connection to an external network connected to said at least one structure.

13

14

15 3. The scalable routing system of claim 1, wherein said at least one
structure includes a crossbar.

16

17

18

19 4. The scalable routing system of claim 1, wherein said at least one
structure includes a bus.

20

21

22

23 5. The scalable routing system of claim 1, wherein said plurality of
data compilers use a hashing function to distribute packet flows among said plurality of
route processing engines.

1

2 6. The scalable routing system of claim 2, wherein both said plurality
3 of data compilers and said at least one uplink connection to an external network use a
4 hashing function to distribute packet flows among said plurality of route processing en-
5 gines.

6

7 7. The scalable routing system of claim 6, wherein said hashing func-
8 tion is designed to maintain the original order of packets in the same flow while allowing
9 packets in different flows to be processed out of order.

10

11 8. The scalable routing system of claim 6, wherein said hashing func-
12 tion is designed to maintain the original order of packets in the same flow while allowing
13 packets in different flows to be processed out of order.

14

15 9. The scalable routing system of claim 8, whereby processing power
16 of said system can be scaled by adding additional route processing engines to said plural-
17 ity of route processing engines.

18

19 10. The scalable routing system of claim 8, whereby the processing
20 power of said system can be scaled by adding additional route processing engines to said
21 plurality of route processing engines.

22

1 11. A scalable routing system for distributing packets in a network, in-

2 cluding

3 a plurality of network interfaces;

4 a plurality of route processing engines;

5 a fabric interconnecting said plurality of network interfaces and said plural-

6 ity of route processing engines;

7 wherein each of said plurality of network interfaces uses a hashing function

8 to distribute packet flows among said plurality of route processing engines; and

9 wherein the processing power of the scalable routing system can be scaled

10 by adding additional route processing engines to said plurality of route processing

11 engines.

12

13 12. The scalable routing system of claim 11, wherein said fabric

14 includes a crossbar.

15

16 13. The scalable routing system of claim 11, wherein said fabric

17 includes a bus.

18

19 14. The scalable routing system of claim 11, wherein said hashing

20 function is designed to maintain the original order of packets in the same flow, while

21 allowing packets in different flows to be processed out of order.

1 15. The scalable routing system of claim 11, wherein said network
2 interfaces include data compilers and port adapters.

3

4 16. The scalable routing system of claim 15, wherein said network
5 interfaces include at least one uplink connection to an external network.

SEARCHED
INDEXED
SERIALIZED
FILED

1 Abstract of the Disclosure

2

3 The invention provides a router architecture that is scalable, that is, as more
4 processing power is desired, more individual processors can be added. The data flow
5 from each line can be distributed among all of the processors in the system. As desired
6 services are added, increasing the amount of “touch” or processing performed on the
7 packets in the system, more processors can be added to carry the increased load. The
8 router architecture is also able to distribute the high interface rate of an uplink connection
9 in the same manner. Packets are allocated to processors in a manner that allows the
10 original order of data packets within the same flow to be maintained. The system uses a
11 hash function to distribute the flows, making sure that packets within the same flow are
12 sent to the same processor so that the original packet order in each flow is maintained.
13 Different flows may be sent to different processors.

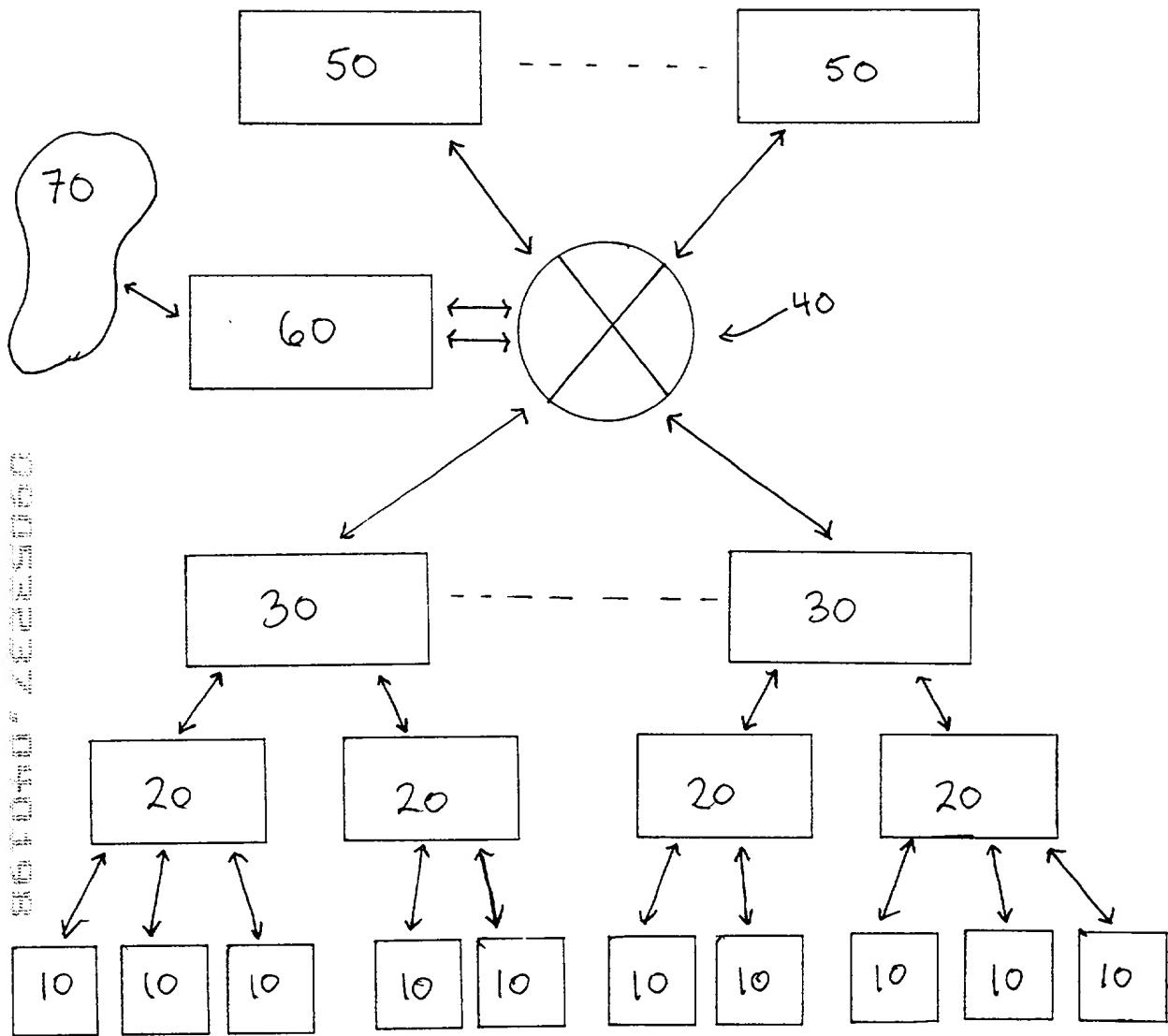


Figure 1 : Router Architecture

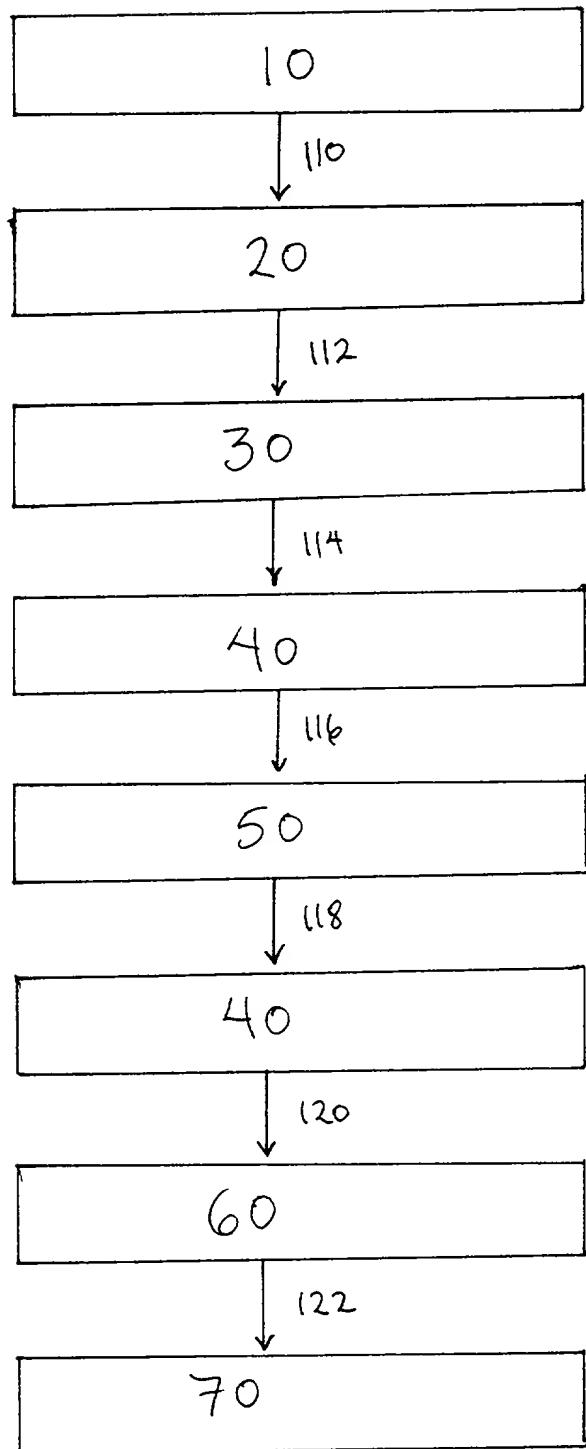


Figure 2: Sample Flow Diagram for a Data Packet

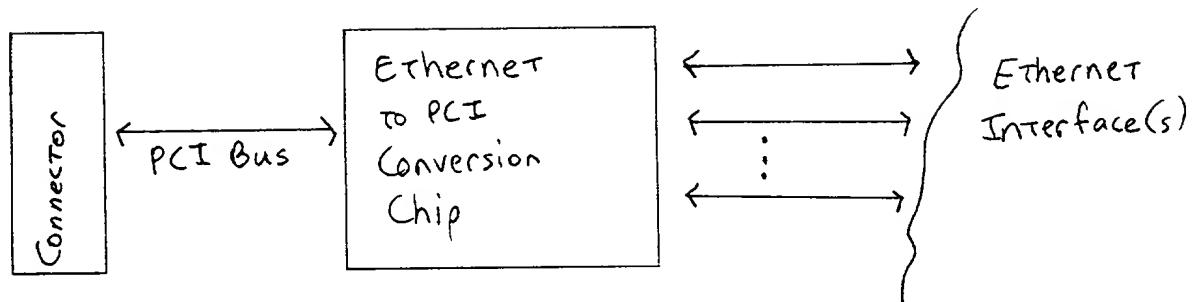


Figure 3a: Port Adaptor Design for an Ethernet Interface

Copyright © 2003, 2004 Pearson Education, Inc.

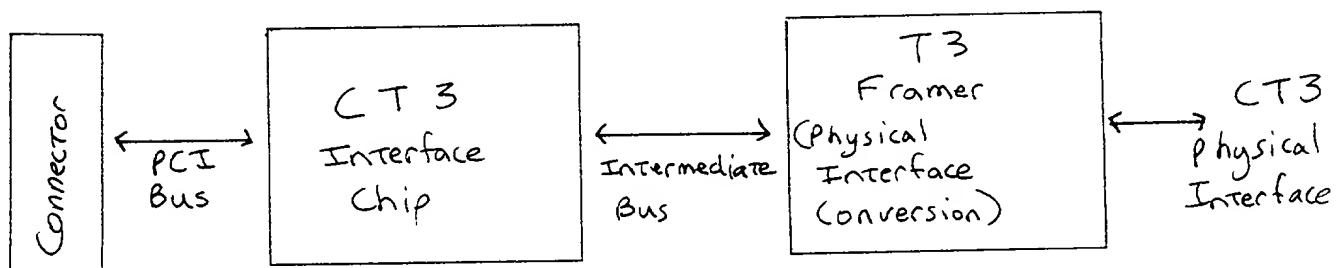


Figure 3b: Port Adaptor Design for a Channelized T3 Line (CT3) Interface

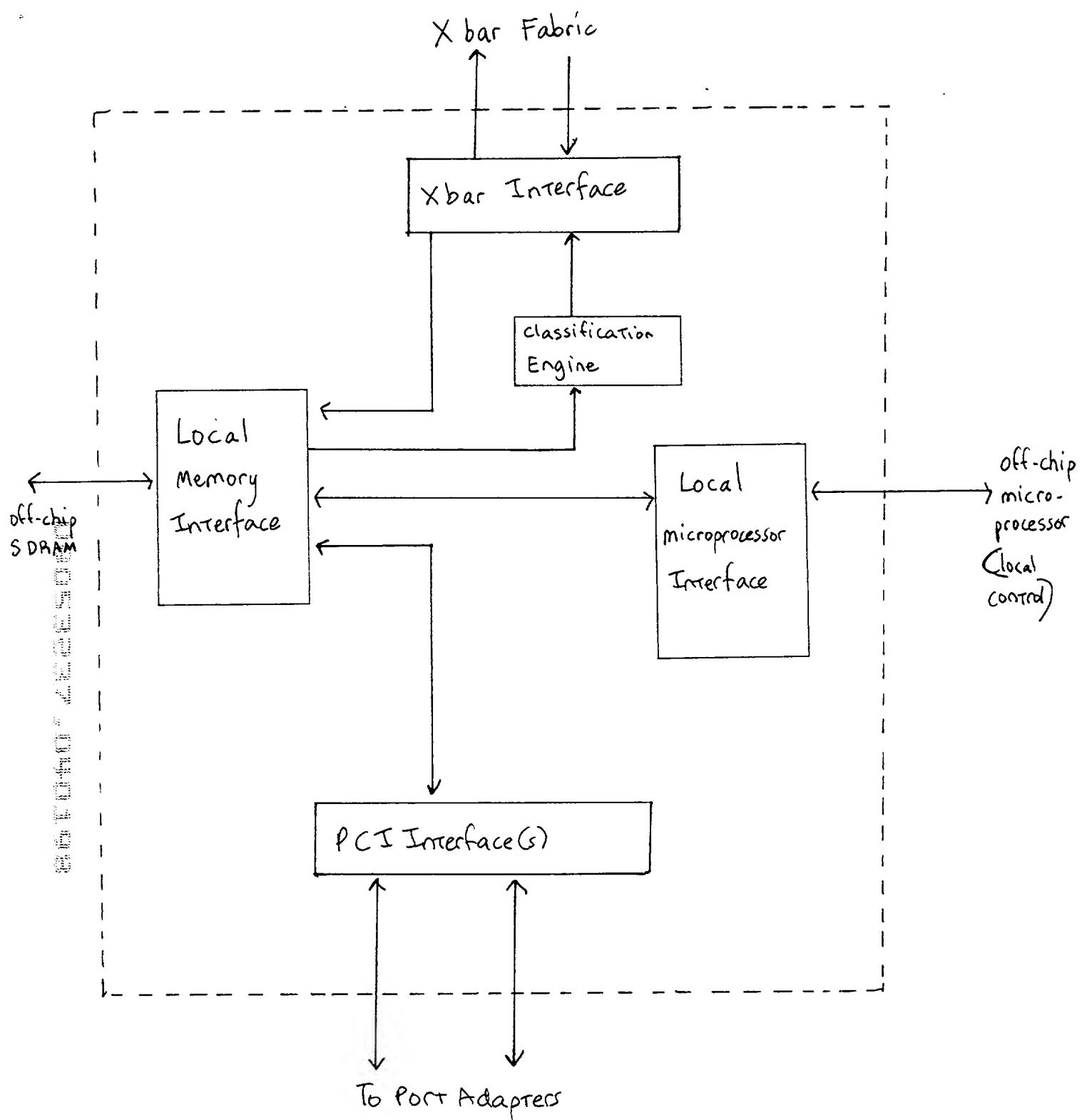
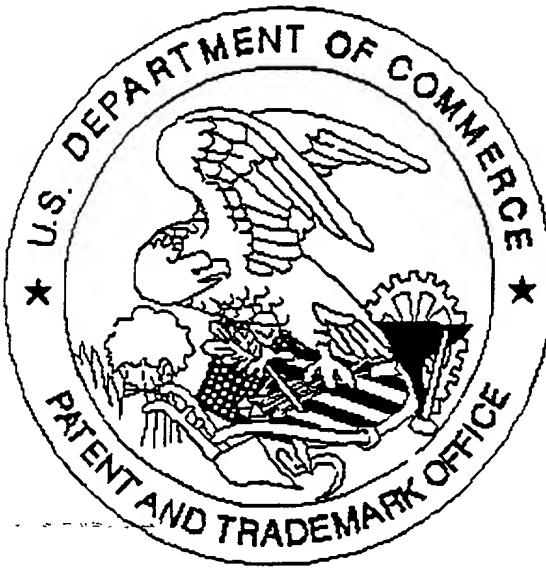


Figure 4: Data Compiler

United States Patent & Trademark Office
Office of Initial Patent Examination -- Scanning Division



Application deficiencies found during scanning:

1. Application papers are not suitable for scanning and are not in compliance with 37 CFR 1.52 because.
 - All sheets must be the same size and either A4 (21 cm x 29.7 cm) or 8-1/2" x 11" Pages _____ do not meet these requirements.
 - Papers are not flexible, strong, smooth, non-shiny, durable, and white.
 - Papers are not typewritten or mechanically printed in permanent ink on one side.
 - Papers contain improper margins. Each sheet must have a left margin of at least 2.5 cm (1") and top, bottom and right margins of at least 2.0 cm (3/4").
 - Papers contain hand lettering.
2. Drawings are not in compliance and were not scanned because:
 - The drawings or copy of drawings are not suitable for electronic reproduction.
 - All drawings sheets are not the same size. Pages must be either A4 (21 cm x 29.7 cm) or 8-1/2" x 11"
 - Each sheet must include a top and left margin of at least 2.5 cm (1"), a right margin of at least 1.5 cm (9/16") and a bottom margin of at least 1.0 cm (3/8").
3. Page(s) _____ are not of sufficient clarity, contrast and quality for electronic reproduction
4. Page(s) _____ are missing.
5. OTHER No Declaration